Technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants

Casacuberta, Josep ; Nogué, Fabien ; Naegeli, Hanspeter ; Birch, Andrew Nicholas ; De Schrijver, Adinda ; Gralak, Mikołaj Antoni ; Guerche, Philippe ; Manachini, Barbara ; Messéan, Antoine ; Nielsen, Elsa Ebbesen ; Robaglia, Christophe ; Rostoks, Nils ; Sweet, Jeremy ; Tebbe, Christoph ; Visioli, Francesco ; Wal, Jean-Michel ; Moxon, Simon ; Schneeberger, Korbinian ; Federici, Silvia ; Ramon, Matthew ; Papadopoulou, Nikoletta ; Jones, Huw

DOI: https://doi.org/10.2903/j.efsa.2018.5345

SCIENTIFIC OPINION

EFSA Journal

ADOPTED: 14 June 2018

doi: 10.2903/j.efsa.2018.5345

Technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants

EFSA Panel on Genetically Modified Organisms (EFSA GMO Panel), Josep Casacuberta, Fabien Nogué, Hanspeter Naegeli, Andrew Nicholas Birch, Adinda De Schrijver, Mikołaj Antoni Gralak, Philippe Guerche, Barbara Manachini, Antoine Messéan, Elsa Ebbesen Nielsen, Christophe Robaglia, Nils Rostoks, Jeremy Sweet, Christoph Tebbe, Francesco Visioli, Jean-Michel Wal, Simon Moxon, Korbinian Schneeberger, Silvia Federici, Matthew Ramon, Nikoletta Papadopoulou and Huw Jones

Abstract

As part of the risk assessment (RA) requirements for genetically modified (GM) plants, according to Regulation (EU) No 503/2013 and the EFSA guidance on the RA of food and feed from GM plants (EFSA GMO Panel, 2011), applicants need to perform a molecular characterisation of the DNA sequences inserted in the GM plant genome. The European Commission has mandated EFSA to develop a technical note to the applicants on, and checking of, the quality of the methodology, analysis and reporting covering complete sequencing of the insert and flanking regions, insertion site analysis of the GM event, and generational stability and integrity. This Technical Note puts together requirements and recommendations for when DNA sequencing is part of the molecular characterisation of GM plants, in particular for the characterisation of the inserted genetic material at each insertion site and flanking regions, the determination of the copy number of all detectable inserts, and the analysis of the genetic stability of the inserts, when addressed by Sanger sequencing or NGS. This document reflects the current knowledge in scientific-technical methods for generating and verifying, in a standardised manner, DNA sequencing data in the context of RA of GM plants. From 1 October 2018, this Technical Note will replace the JRC guideline of 2016 (updated April 2017) related to the verification and quality assessment of the sequencing of the insert(s) and flanking regions. It does not take into consideration the verification and validation of the detection method which remains under the remit of the JRC.

Keywords: DNA sequencing, next-generation sequencing, genetically modified organisms, molecular characterisation, risk assessment, NGS, genetic stability

Requestor: European Commission

Question number: EFSA-Q-2017-00706

Correspondence: gmo@efsa.europa.eu

GMO Panel members:. Hanspeter Naegeli, Andrew Nicholas Birch, Josep Casacuberta, Adinda De Schrijver, Mikołaj Antoni Gralak, Philippe Guerche, Huw Jones, Barbara Manachini, Antoine Messéan, Elsa Ebbesen Nielsen, Fabien Nogué, Christophe Robaglia, Nils Rostoks, Jeremy Sweet, Christoph Tebbe, Francesco Visioli and Jean-Michel Wal.

Acknowledgements: The GMO Panel wishes to thank the following for the support provided to this scientific output: Cristian Savini, Mauro Petrillo and Hans Moons.

Suggested citation: EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), Casacuberta J, Nogué F, Naegeli H, Birch AN, De Schrijver A, Gralak MA, Guerche P, Manachini B, Messéan A, Nielsen EE, Robaglia C, Rostoks N, Sweet J, Tebbe C, Visioli F, Wal J-M, Moxon S, Schneeberger K, Federici S, Ramon M, Papadopoulou N and Jones H, 2018. Scientific Opinion on the technical Note on the quality of DNA sequencing for the molecular characterisation of genetically modified plants. EFSA Journal 2018;16(7):5345, 11 pp. https://doi.org/10.2903/j.efsa.2018.5345

ISSN: 1831-4732

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

Summary

Genetically modified organisms (GMOs) are subject to a risk assessment (RA) and regulatory approval before entering the European market. In this process, the role of the European Food Safety Authority (EFSA) is to independently assess and provide scientific advice to risk managers on any possible risk that the use of GMOs may pose to human and animal health and the environment. As part of the RA requirements for genetically modified (GM) plants, according to Regulation (EU) No 503/2013 and the EFSA guidance on the RA of food and feed from GM plants (EFSA GMO Panel, 2011), applicants need to perform a molecular characterisation of the DNA sequences inserted in the GM plant genome.

The European Commission has mandated EFSA to develop a technical note to the applicants on, and checking of, the quality of the methodology, analysis and reporting covering complete sequencing of the insert and flanking regions, insertion site analysis of the GM event, and generational stability and integrity. This Technical Note puts together requirements and recommendations for when DNA sequencing is part of the molecular characterisation of GM plants, in particular for the characterisation of the inserted genetic material at each insertion site and flanking regions, the determination of the copy number of all detectable inserts, and the analysis of the genetic stability of the inserts, when addressed by Sanger sequencing or NGS.

This document reflects the current knowledge in scientific-technical methods for generating and verifying, in a standardised manner, DNA sequencing data in the context of RA of GM plants. This Technical Note takes as a starting point the JRC guideline 2016 (updated April 2017)[1] on the quality and reliability of submitted information related to sequencing of the insert(s) and flanking regions, integrating and updating the JRC guideline where scientifically justified. From 1 October 2018, this Technical Note will replace the JRC guideline of 2016 (updated April 2017) related to the verification and quality assessment of the sequencing of the insert(s) and flanking regions. It does not take into consideration the verification and validation of the detection method which remains under the remit of the JRC.

A list of information that should be included in GMO applications submitted to EFSA in conjunction with the DNA sequences can be found in Annex 1. In order to assist in the submission of sequencing information in accordance with this Technical Note, and to enhance the efficiency of the compliance checks, applicants are requested to implement a harmonised structure of such information and data as described in Annex 2.

[1] EURL-JRC Guideline for the submission of DNA sequences derived from genetically modified organisms and associated annotations within the framework of Directive 2001/18/EC and Regulation (EC) No 1829/2003, European Union, 2016.

Table of contents

1. Introduction

Genetically modified organisms (GMOs) are subject to a risk assessment (RA) and regulatory approval before entering the European market. In this process, the role of the European Food Safety Authority (EFSA) is to independently assess and provide scientific advice to risk managers on any possible risk that the use of GMOs may pose to human and animal health and the environment. As part of the RA requirements for genetically modified (GM) plants, according to Regulation (EU) No 503/2013[2] and the EFSA guidance on the RA of food and feed from GM plants (EFSA GMO Panel, 2011), applicants need to perform a molecular characterisation of the DNA sequences inserted in the GM plant genome.

At the nucleic acid level, the molecular characterisation for the RA of GM plants includes among other analyses, the following three aspects that could be addressed by DNA sequencing: (1) the determination of the copy number of all detectable inserts, both complete and partial; (2) the determination of the organisation and sequence of the inserted genetic material at each insertion site, as well as that of the 5' and 3' flanking regions hereafter referred to as the characterisation of the insert and flanking regions; and (3) the analysis of the genetic stability of the inserts.

Usually the characterisation of the sequence of the insert is performed by Sanger sequencing, while the copy number of detectable insert(s) and the genetic stability of the plant insertion sites are determined by Southern analysis. Recently, next-generation sequencing (NGS) has been used in biological research to allow for rapid, high-throughput nucleic acid sequencing in a variety of applications. This fast-evolving field includes different techniques and can be used as an alternative approach for the molecular characterisation of GM plants in the frame of RA aspects. NGS technologies can address the aspects of the molecular characterisation of GM plants mentioned above and offer an alternative to Southern analysis, as junction sequence analysis (JSA) can be performed starting from NGS data (Kovalic et al., 2012; Yang et al., 2013; Pauwels et al., 2015; Guo et al., 2016; Guttikonda et al., 2016).

This Technical Note puts together requirements and recommendations for the quality of DNA sequencing, based on Sanger sequencing or NGS technologies, that can be used for parts of the molecular characterisation of GM plants in the context of RA. To ensure that the quality parameters used for the sequencing methodologies are in line with up-to-date scientific knowledge, as the technologies advance, this Technical Note will be updated when needed.

A list of information that should be included in GMO applications submitted to EFSA in conjunction with the DNA sequences can be found in Annex 1 (see Supporting information). In order to assist in the submission of sequencing information in accordance with this Technical Note, and to enhance the efficiency of the compliance checks, applicants are requested to implement a harmonised structure of such information and data as described in Annex 2 (see Supporting information).

1.1. Terms of Reference as provided by the European Commission

The Commission mandated EFSA to develop *a technical note to the applicants on, and checking of, the quality of the methodology, analysis and reporting covering complete sequencing of the insert and flanking regions, insertion site analysis of the genetically modified (GM) event, and generational stability and integrity*. This technical note takes as a starting point the JRC guideline 2016 (updated April 2017)[1] on the quality and reliability of submitted information related to sequencing of the insert(s) and flanking regions, integrating and updating the JRC guideline where scientifically justified. In accordance with the mandate of the EC, EFSA will take over the verification and quality assessment of the sequencing data from the JRC for all GMO applications received after 1 October 2018. For all applications received after October 1, 2018, this document replaces the JRC guideline of 2016 (updated April 2017).[1] This document does not take into consideration the verification and validation of the detection method, which remains under the remit of the JRC. Verification and quality assessment of the sequencing information in GMO applications submitted before 1 October 2018 will be handled by the JRC according to the EURL-JRC guidance.[1]

2. Data and methodologies

To address this mandate, EFSA established a GMO Panel working group consisting of EFSA staff and experts specialised in the field. JRC representatives were invited as observers to this working group. In delivering this Technical Note, EFSA took into account the requirements of the Guidance on the RA of

[2] Commission Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. OJ L157, 8.6.2013, p. 1–48.

food and feed from GM plants (EFSA GMO Panel, 2011) and of Regulation (EU) No 503/2013. This document applies when sequencing is used for the determination of the copy number of all detectable inserts, the characterisation of the inserted genetic material at each insertion site and flanking regions, and the analysis of the genetic stability of the inserts.

This document takes into account the current knowledge in scientific-technical methods for generating and verifying in a standardised manner, DNA sequencing data in the context of RA of GM plants. Data from published scientific literature, and experience from GMO applications containing data sets generated by Sanger sequencing or NGS were considered, in order to produce detailed recommendations on Sanger and NGS-generated data sets.

3. Requirements for the material and DNA sample preparation

The material used for DNA sample preparation should be derived from the GM plant to be assessed. In case of stacks, the material should come from the GM plant (containing all events) under assessment. The applicant should provide a report clearly describing the source of the plant material specifically indicating the GM event(s) in the GM plant, with the unique identifier corresponding to the GMO, the plant species and the generation in the breeding tree. The applicant should also include a description on how and on what date the plant material was collected, specify the organ and/or tissues as well as the number of plants from which the DNA sample(s) used for sequencing was prepared. The DNA extraction protocol should be included in the report. If multiple DNA extractions are needed, it is strongly recommended to use the same DNA extraction protocol for sample preparation.

It is recommended, that an amount of sample sufficient for at least three further sequencing experiments, i.e. activities that lead to the generation of a final event sequence, should be stored from the submission of the application in case reanalysis is requested.

4. Requirements for the sequencing quality, specific to the technology used

This section provides requirements and recommendations on the information to be submitted in GMO applications when DNA sequencing approaches have been used for any of the molecular characterisation aspects that could be addressed by DNA sequencing. The two main technologies that are currently used in the context of RA of GM plants are Sanger sequencing and NGS.

4.1. Sanger sequencing

This section provides the general requirements and recommendations on the information to be submitted when Sanger sequencing is used in GMO applications.

For the characterisation of the event(s), the final sequence submitted for each event (hereafter referred to as final event sequence) should be generated from at least two independent polymerase chain reactions (PCRs) covering every position of the sequence. The sequence should be produced by bidirectional sequencing, i.e. each base should be sequenced on the two DNA strands. Since at least two independent sequencing experiments are required, the raw sequence of each nucleotide should be covered at least four times.

The applicant should provide a report describing, as a minimum, the overall strategy to obtain the DNA fragment(s) (e.g. subcloning, long-run PCR) used for sequencing, the sequencing strategy and the details of the methodology and experimental design used to obtain the final event sequence.

The applicant should submit all individual sequences, alignments and final event sequence(s) for the GM plant under assessment as described in Section 6. Any uncertainty observed in the raw data and any manual editing performed on the sequence (base calling and trimming) should be reported and justified.

4.2. Next Generation Sequencing

This section provides requirements when NGS is used in GMO applications and describes the most relevant parameters to be considered when NGS methodology and generated data sets are assessed in applications. A report on the sequencing strategy and the details of the experiment has to be provided for the final sequence(s) submitted for the event(s). This report should include, at least, the description of the technology used, the sequencing method and the details of the experimental design. The applicant should submit all sequences, alignments and final event(s) sequence(s) as described in Section 6.

4.2.1. Quality of data sets

In order to assess NGS data sets, information on sequencing platforms used to generate the data together with the number and quality statistics of reads generated for each experiment should be described and included. Providing this information is especially important when reads are not aligned to a reference genome, as such alignments would allow for an additional estimation of quality. Common tools providing quality statistics of reads, like average base quality across the reads and flagging potential quality issues, including over-represented reads (or k-mers) and contamination, are available and should be used. For example, FASTQC is a widely used tool for checking the quality of NGS data sets (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

The applicant should provide raw read numbers for each sequencing run. As trimming and quality filtering are often used to remove poor quality and low complexity reads and to trim sequencing adaptors or low quality ends, the strategy for sequence trimming, the number of reads discarded and trimmed as well as the average read lengths after each step of filtering and trimming, should all be discussed in the application to allow assessment of the methodology.

4.2.2. Library preparation and sequencing strategy

Information on the library construction method has to be provided. A detailed description of how each of the sequencing libraries was prepared along with details of the sequencing chemistry, strategy and platform should be given. In addition, if a sequence capture approach is used, it is critical that applicants thoroughly describe all experimental procedures and probe design, as well as how hybridisation conditions and capture efficiency were assessed.

4.2.3. Read depth

Different sequencing technologies producing reads of variable quality and length are currently available. The number of reads that cover a particular position (read depth) needed to obtain the final event sequence depends, among other factors, on the quality, the length of the individual reads and the purpose of the sequencing experiment. In order to assess the NGS data submitted by the applicant, information on the read depth, and where relevant (e.g. for whole-genome sequencing (WGS)) the average read depth and its variation, should be provided. The applicant should justify the (average) read depth based on the methodology and technology used.

Average read depth description when using WGS for insert identification

When WGS is used for the identification of the insert(s) and the possible insertion(s) of backbone sequences, there is a need to estimate the average read depth across the whole genome. To estimate the average read depth, different approaches are possible. For example, in cases where a reference genome is available, reads should be aligned to the entire sequence to calculate average read depth. In cases where this is not possible or favourable, reads should be aligned to several reference genes/genomic regions from diverse genomic locations to assess read depth locally. In cases where no genomic resources exist, theoretical average read depth metrics for WGS data, derived from the equation discussed below, should be used. Applicants should have a good estimate of the genome size of the GM plant and therefore should be able to calculate the number of reads required to cover the genome to a specified depth. This can be achieved using the Lander–Waterman formula (Lander and Waterman, 1988), where average read depth is indicated as coverage:

$$\text{Coverage (average read depth)} = \frac{\text{number of reads} \times \text{read length}}{\text{estimated genome size}}$$

The Lander–Waterman equation gives a theoretical average read depth. However, this equation does not consider platform- and sequence-specific biases (Ross et al., 2013) and provides estimate of the average read depth, which is a limitation as the read depth is not necessarily uniform across the genome (Sims et al., 2014). The applicant should also consider evaluating the number of reads corresponding to mitochondrial or plastid DNA and justify the read depth of the nuclear DNA, since the technology used and the genome of the respective GM plant may affect the average read depth calculation (Lutz et al., 2011).

Minimum read depth for insert description

As already commented, the minimum read depth will depend on different factors, including the approach used. For example, at the current state of NGS technologies, when short read technologies are used (e.g. those currently marketed by Illumina) for the sequencing of the insert(s) and flanking regions the minimum read depth should not be below 40.

4.2.4. Description of bioinformatic analysis

The applicant can choose any appropriate bioinformatics pipeline for the various analyses of NGS data sets; however, the methodology and tools used should be thoroughly described by the applicant. In particular for unpublished or in-house tools, a full description along with the scripts, source code and pipelines, inputs and outputs of each of the steps in the analysis, and other parameters used should be provided. Any filtering of results and thresholds should be described and justified. In addition, a flowchart of the analysis showing how the raw data were processed, from start to end, in order to obtain the final results should be formulated and submitted for each final event sequence (Ekblom and Wolf, 2014). When common bioinformatics software such as BLAST and common tools for read filtering and trimming are used, the exact tool version must be provided. Since each tool includes multiple parameters and options, the exact parameters and options applied should be specified and justified in order to flag potential issues and ensure transparency.

5. Additional considerations for the sequencing quality, specific to the molecular characterisation aspects

This section describes considerations and requirements, in addition to those described in Sections 4 and 6, on the information to be submitted in GMO applications for each of the three specific aspects of the molecular characterisation that could be addressed by one or a combination of the DNA sequencing approaches described in Section 4.

5.1. Sequencing for the characterisation of the insert(s) and flanking regions

In order to risk assess a GM event, the applicant has to characterise the sequence of the insert(s) and genomic flanking regions (EFSA GMO Panel, 2011, and Regulation (EU) No 503/20131)[2]. The final event sequence and alignments have to be submitted as described in Section 6.

In cases where the applicant has previously submitted the sequence of an event to the European Commission, EFSA or the EURL-GMFF, they are required to compare the sequence of the GM event under assessment with all previously submitted sequences of this event. The applicant has to provide an alignment including all those sequences and report the differences found and discuss the reasons for the differences.

5.1.1. Considerations when Sanger sequencing is used

When Sanger sequencing is used for the characterisation of the insert and genomic flanking regions, the applicant should comply with the requirements discussed in Sections 3 and 4.1.

5.1.2. Considerations when NGS is used

For the determination of the insert(s) sequence and genomic flanking regions, as an alternative to Sanger sequencing, different NGS approaches may be used, such as WGS, or sequence capture approaches to enrich for the target DNA fragments before sequencing (Ekblom and Wolf, 2014; Inagaki et al., 2015). Although in some cases, this can be relatively straightforward, some configurations of the inserted sequences can make this more challenging, e.g. the presence of sequence rearrangements or duplications within the locus, or the nature of the inserted sequence, including the presence of long repeats. A combination of approaches, including ultralong reads, sequencing of cloned genomic fragments or PCR amplicons (including by Sanger sequencing) may be needed in such cases. The applicant is required to describe, discuss and justify the approach used. In particular, discussion and justification on read depth has to be provided according to the recommendations for minimum read depth in Section 4.2.3. Regardless of the NGS approach used, the

sequence of the insert(s) and genomic flanking regions, and alignments, should be provided following the requirements described in section 6.

5.2. Determining the copy number of all detectable inserts

The determination of the copy number of all detectable inserts is required as part of the molecular characterisation of the GM plant. This can be achieved in a number of ways including junction sequence analysis (for example, see Kovalic et al., 2012). This approach relies on the computational identification of junction reads that show both sequence identity with the insert or the vector sequence and with the host genome (*chimeric* reads). Because these reads have a partial match to both the insert/vector and the host genome, reads of sufficient length (approx. 100 bp) are required to accurately identify junctions. Any discarding of possible junction reads should be described and justified as described in Section 4.2.1.

Read depth for junction sequence analysis

As discussed in Section 4.2.3, the read depth is a key factor to evaluate the quality of the data. The applicant should include detailed information on (average) read depth, as described in Section 4.2.3. Although this is dependent on the characteristics of the genome and the sequencing technology used, read depth should be sufficiently high to detect junction reads, and justified by the applicant. Willems et al. (2016) have proposed a statistical framework for estimating the probability of sequencing junction reads that span the junction between the intended introduced DNA and the host genome ('identification approach') which may be useful for the applicant to consider when planning such experiments. A combination of approaches could also be used.

5.3. Genetic stability

In the case of GM plants containing a single event, genetic stability encompassing (a) the Mendelian inheritance of the insert(s) and (b) the molecular stability of the event over several generations, has to be demonstrated. The Mendelian inheritance is currently checked by segregation analysis and the Chi-square test. Molecular stability of the event over several generations has usually been demonstrated by Southern experiments or PCR. Different sequencing approaches could be an alternative to the currently used methods to demonstrate that the insertion site(s) and the structure of the insert(s) is maintained over several generations. This can be accomplished by mapping of NGS reads (or contigs) to the sequence of the insert(s) and the flanking regions.

In the case of GM plants containing multiple events, the integrity of each event in the stack should be demonstrated. When sequencing by Sanger sequencing or NGS is done, this should be performed following the requirements and recommendations in Section 4.

The final event sequence(s) should be submitted as described in Section 6.

6. Data format requirements[3]

The raw data should be provided in specific formats depending on the methodology used, as indicated in the following paragraphs.

The final event sequence has to be submitted as electronic ASCII text files using either the EMBL/GenBank format, or, preferably, the NCBI's Sequin (ASN.1) format and shall be annotated according to the INSDC Feature Table Definition Document[2] with at least the following descriptors and features, including their location on the sequence:

- "DEFINITION" (Title describing the sequence record)
- "SOURCE", "ORGANISM" (according to the NCBI Taxonomy database)
- "SIZE" (in base pairs)
- "MOLECULE TYPE" (DNA)
- "TOPOLOGY" (linear/circular),
- "REFERENCE" (References with Authors, Title, Journals, etc.)
- "Source" (regions/sources of GMO insert and host organism)

[3] The data format requirements apply also to the sequence information needed for the verification of the detection method. In addition, for the detection method verification only, if the sequence of a taxon-specific reference gene is included in the submission, the full sequence of the taxon-specific target and its GenBank accession number shall also be submitted (see also EURL-JRC Technical report on the Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing (see http://gmo-crl.jrc.ec.europa.eu/guidancedocs.htm)

- "STS" (Sequence Tagged Site corresponding to the PCR amplicon of the detection method)
- "Primer bind" (with primer name and sequence for Forward, Reverse Primer and Probe)
- All genetic elements ("gene", "promoter", "terminator", etc.)
- All coding sequences ("CDS"), including their translation.

6.1. Data format requirements for Sanger experiments

The applicant should submit all individual sequences of each event in ABI or FASTQ format. These sequences should be aligned to generate a final event sequence and the alignment should be submitted in CLUSTAL or FASTA format. The final event sequence should be submitted as described above.

6.2. Data format requirements for NGS experiments

The application should include raw NGS reads in compressed (such as gzip) FASTQ format. The sequences aligned/mapped to and used to generate the final event sequence should be provided in Sequence Alignment/Map (SAM) format (Li et al., 2009), Binary Alignment/Map (BAM) format (Li et al., 2009) or CRAM[4] format. It is also suggested to submit an additional ACE[5] file. The final event sequence should be submitted as described above.

Information required

A list of information that should be included in GMO applications submitted to EFSA in conjunction with the DNA sequences can be found in Annex 1 (in Supporting information). This reporting form should be filled in, signed and provided by the applicant together with the sequencing information.

In order to assist in the submission of sequencing information in accordance with this Technical Note, and to enhance the efficiency of the compliance checks, applicants are requested to implement a harmonised structure of such information and data as described in Annex 2 (in Supporting information).

Documentation provided to EFSA

1) Letter from European Commission to EFSA received on 4 October 2017 for the development of a technical note on and checking of the quality of the methodology, analysis and reporting covering full sequencing and insertion site analysis of the genetically modified event, and generational stability and integrity.
2) Acknowledgement letter dated 16 October 2017 from EFSA to European Commission.
3) Document from EuropaBio to EFSA dated 16 February 2018 on Best Practices for Use of Next Generation Sequencing Methods for the Characterisation of GM Crops.
4) Letter from EuropaBio to EFSA dated 17 April 2018 providing clarifications.

References

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2011. Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011; 9(5):2150, 37 pp. https://doi.org/10.2903/j.efsa.2011.2150

Ekblom R and Wolf JBW, 2014. A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications, 7, 1026–1042. https://doi.org/10.1111/eva.12178

Guo B, Guo Y, Hong H and Qiu LJ, 2016. Identification of genomic insertion and flanking sequence of G2-EPSPS and GAT transgenes in soybean using whole genome sequencing method. Frontiers in Plant Science, 7, 1009, https://doi.org/10.3389/fpls.2016.01009

Guttikonda SK, Marri P, Mammadov J, Ye L, Soe K, Richey K, Cruse J, Zhuang M, Gao Z, Evans C, Rounsley S, 2016. Molecular characterisation of transgenic events using next generation sequencing approach. PLoS ONE, 11, e0149515. https://doi.org/10.1371/journal.pone.0149515

Inagaki S, Henry IM, Lieberman MC and Comai L, 2015. High-throughput analysis of T-DNA location and structure using sequence capture. PLoS ONE, 10, e0139672. https://doi.org/10.1371/journal.pone.0139672

Kovalic D, Garnaat C, Guo L, Yan Y, Groat J, Silvanovich A, Ralston L, Huang M, Tian Q, Christian A, Cheikh N, 2012. The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterisation of crops improved through modern biotechnology. The Plant Genome, 5, 149–163. https://doi.org/10.3835/plantgenome2012.10.0026

[4] https://www.ebi.ac.uk/ena/software/cram-toolkit
[5] https://academic.oup.com/bioinformatics/article/31/19/3216/211288

Lander ES and Waterman MS, 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics, 2, 231–239.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lutz KA, Wang W, Zdepski A and Michael TP, 2011. Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. BMC Biotechnology, 11, 54. https://doi.org/10.1186/1472-6750-11-54

Pauwels K, De Keersmaecker SC, De Schrijver A, du Jardin P, Roosens NH, Herman P, 2015. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: Added value or not? Trends in Food Science & Technology, 45, 319–326. https://doi.org/10.1016/j.tifs.2015.07.009

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB, 2013. Characterizing and measuring bias in sequence data. Genome Biology, 14, R51. https://doi.org/10.1186/gb-2013-14-5-r51

Sims D, Sudbery I, Ilott NE, Heger A and Ponting CP, 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews Genetics, 15, 121–132. https://doi.org/10.1038/nrg3642

Willems S, Fraiture MA, Deforce D, De Keersmaecker SC, De Loose M, Ruttink T, Herman P, Van Nieuwerburgh F, Roosens N, 2016. Statistical framework for detection of genetically modified organisms based on Next Generation Sequencing. Food Chemistry, 192, 788–798. https://doi.org/10.1016/j.foodchem.2015.07.074

Yang L, Wang C, Holst-Jensen A, Morisset D, Lin Y, Zhang D, 2013. Characterisation of GM events by insert knowledge adapted re-sequencing approaches. Scientific Reports, 3, 2839. https://doi.org/10.1038/srep02839

Abbreviations

BAM	Binary Alignment/Map
EURL-GMFF	European Reference Laboratory- Genetically modified food and feed
GM	genetically modified
GMO	genetically modified organisms
JRC	Joint Research Centre
JSA	junction sequence analysis
NGS	next-generation sequencing
PCR	polymerase chain reaction
RA	risk assessment
SAM	Sequence Alignment/Map
WGS	whole-genome sequencing